National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Steering the Ship

Making Sense of Multi Container Deployments with the Help of Kubernetes and AWS

Presented By: Frank Greguska

Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive, Pasadena, CA 91109-8099, U.S.A.

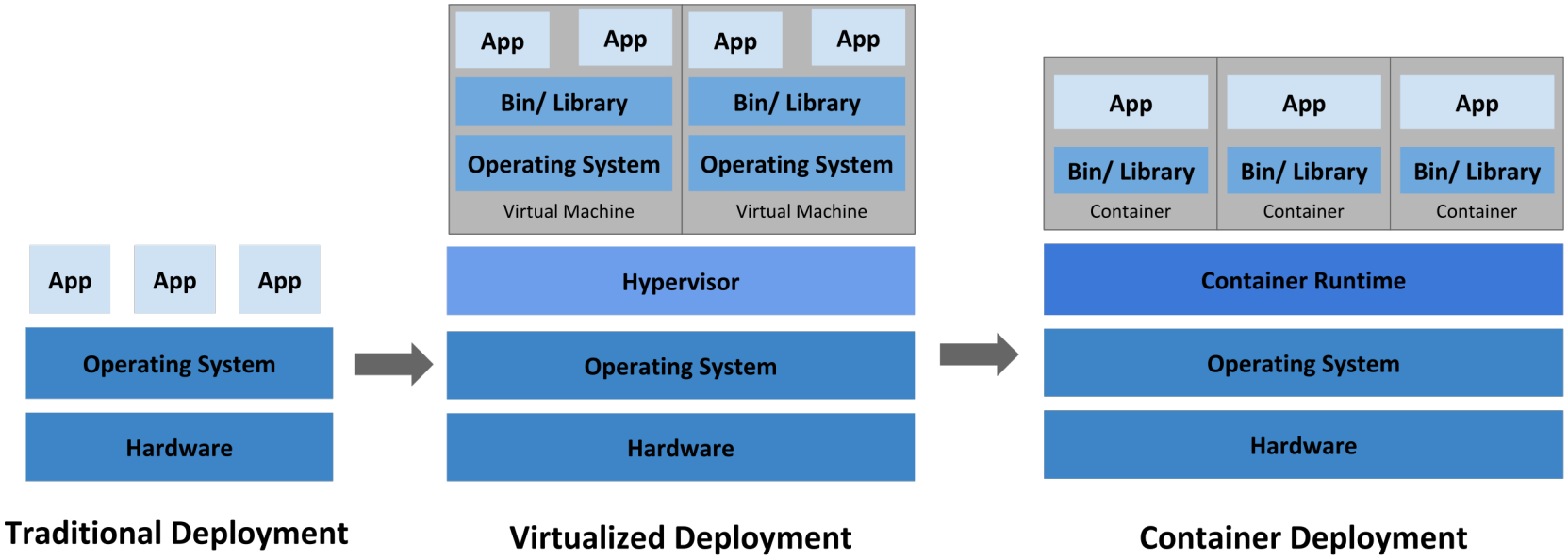
CL#xx-xxxx

Overview

- Why Kubernetes
- Kubernetes Crash Course
- Tales of a Production Deployment using EKS and NASA Sea Level Change Portal

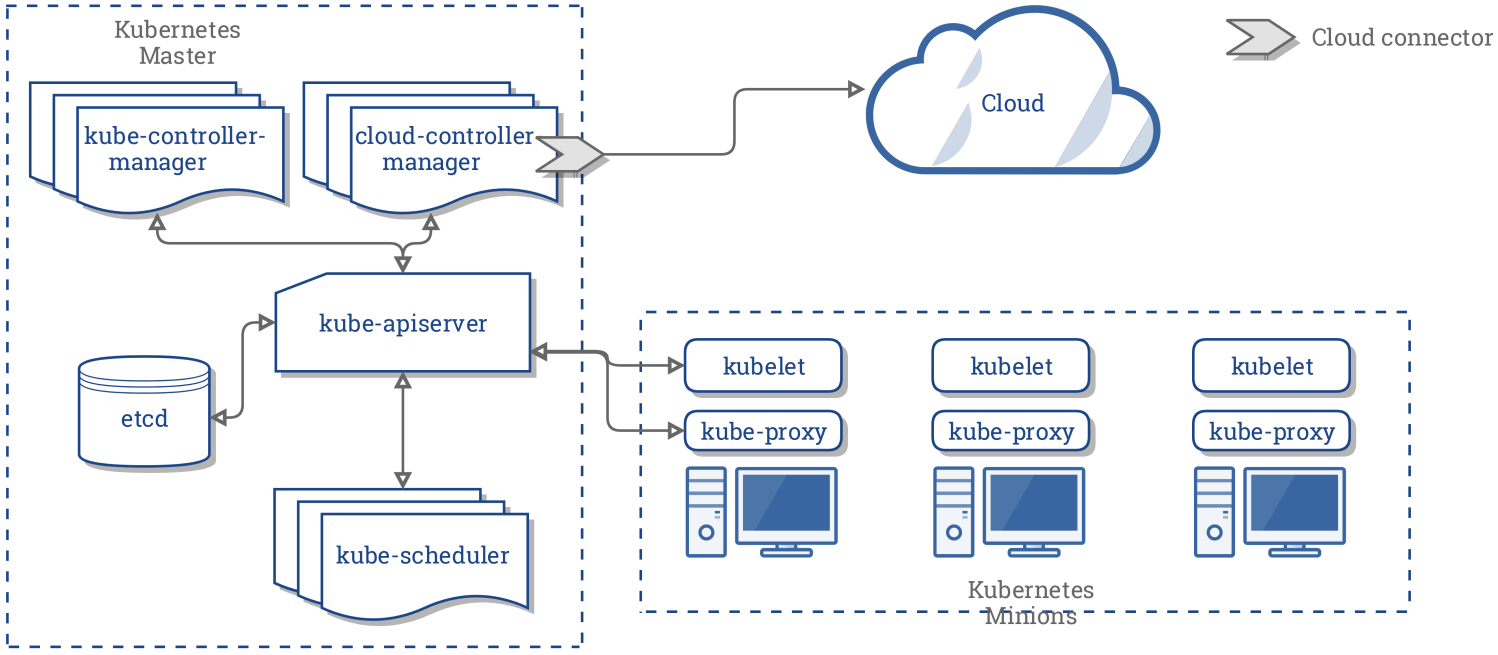
First Comes Docker

- Containers everywhere!
- Basically acts as a Virtual Machine
- Great for encapsulating runtime for a single process/application
- But what about multiple processes/applications?
- Docker Compose!



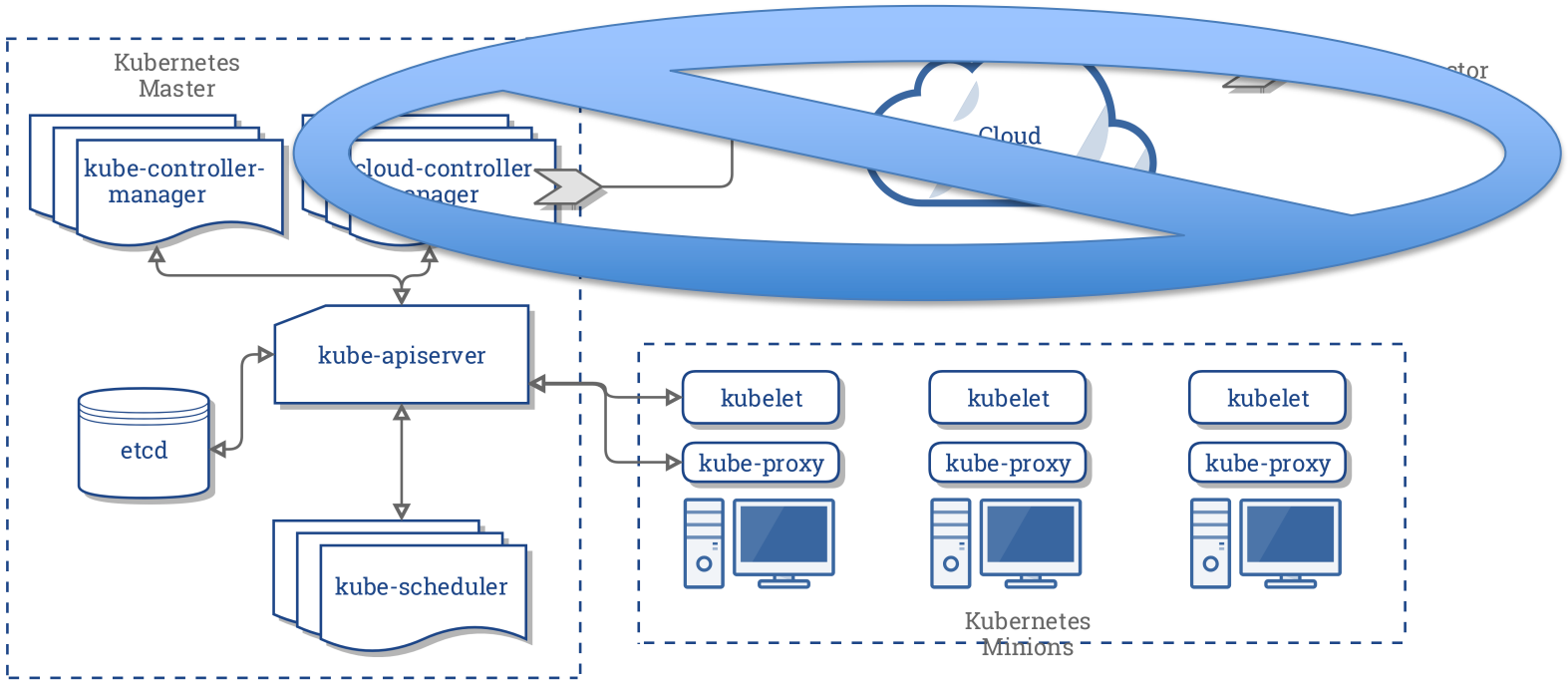
Then Comes Orchestration

- But wait… Compose only launches containers on one host OS
- What if I want to scale across many host OS?
- Kubernetes!



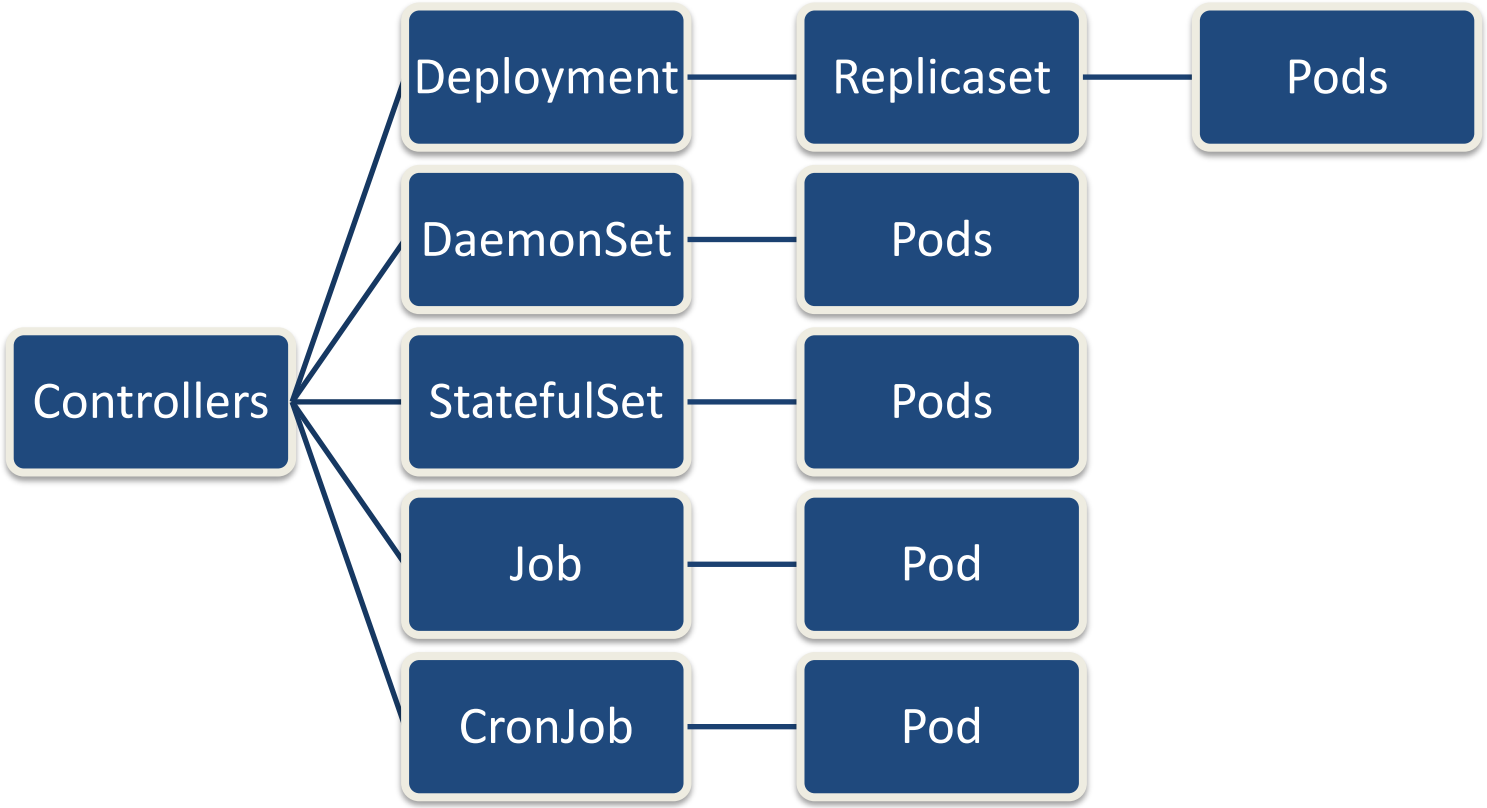
Then Comes Orchestration

- But wait… Compose only launches containers on one host OS
- What if I want to scale across many host OS?
- Kubernetes!



Kubernetes to the Rescue

- Service discovery and load balancing
- Storage orchestration
- Automated rollouts and rollbacks
- Automatic bin packing
- Self healing
- Secret and configuration management



https://towardsdatascience.com/key-kubernetes-concepts-62939f4bc08e

To The Cloud

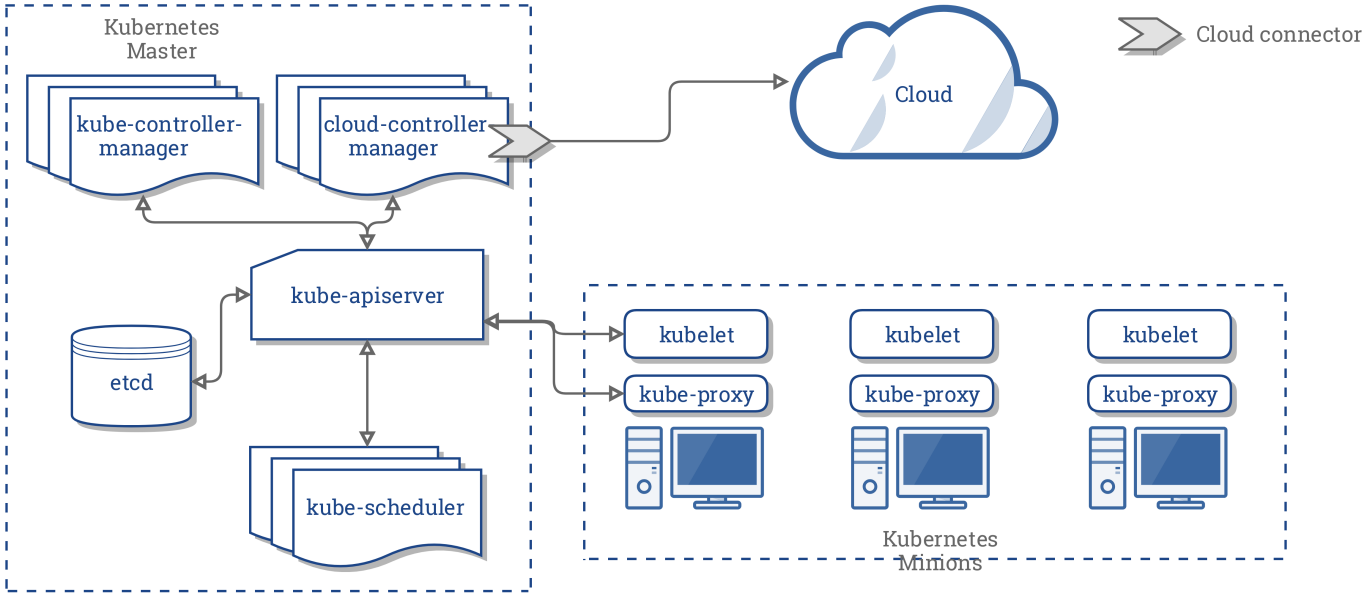
- Kubernetes was built to run on premise or on the cloud
- Cloud controller manager encapsulates cloud specific concerns
- Many hosted offerings available

https://landscape.cncf.io/

Amazon EKS

Google Kubernetes Engine

Azure Kubernetes Service (AKS)



Planning a Production Ready Cluster

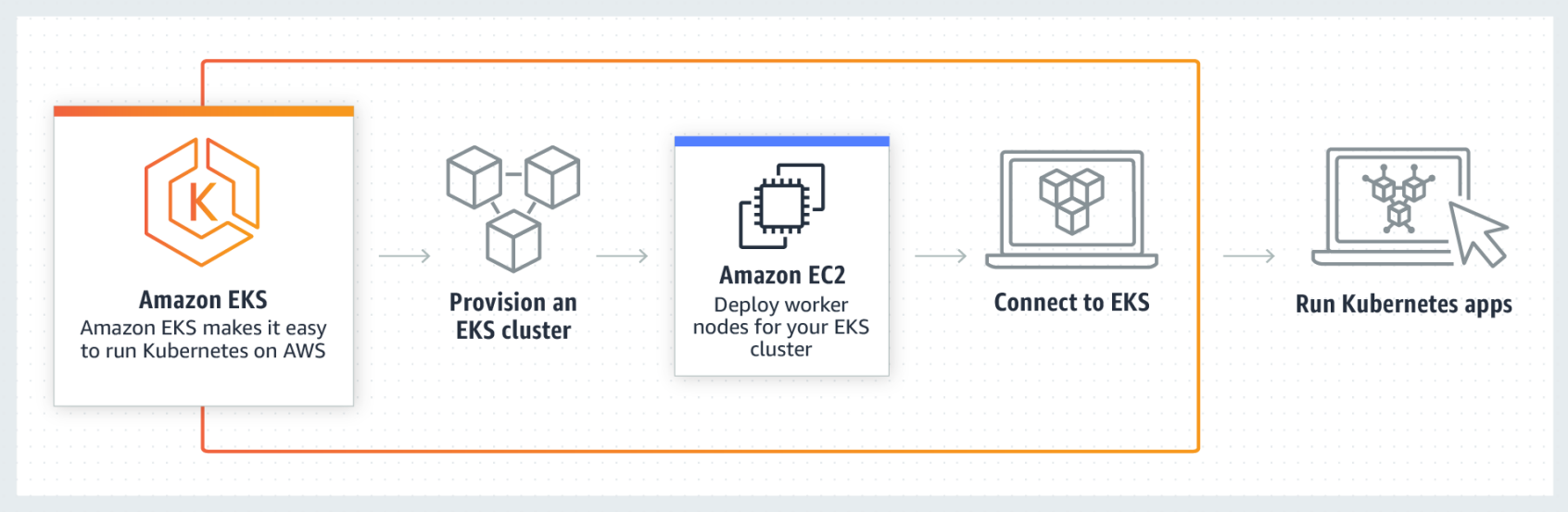
Ship It!

Getting Down to Brass Tacks

- Network Isolation
- Separation of Environments
- Authentication and Authorization
- Resource Quotas
- Scaling
- Day 2 Operations
- *(Cost)*

Everyone Loves Amazon Web Services (AWS)

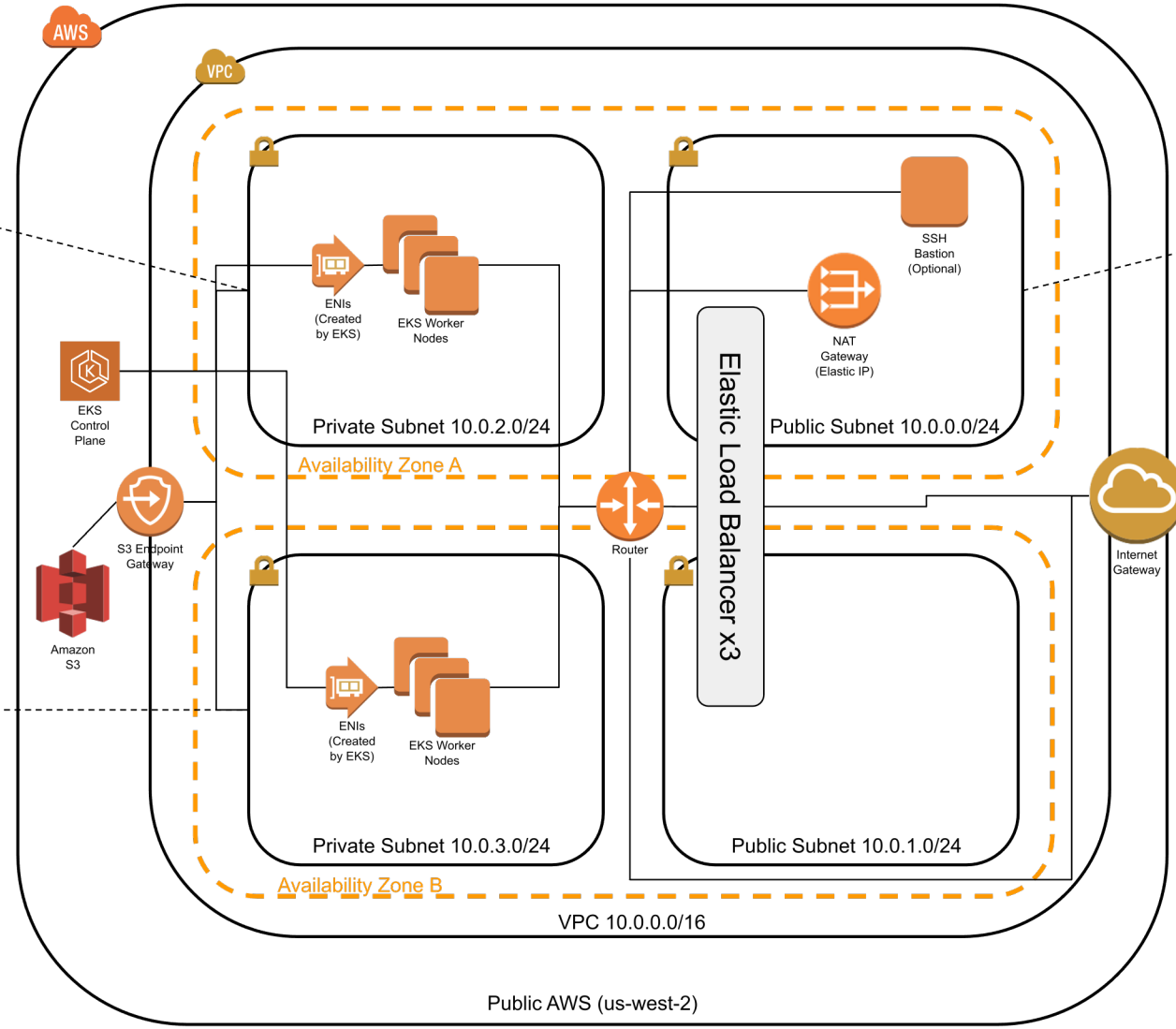
- Amazon Elastic Kubernetes Service (EKS) is a service from Amazon that hosts Kubernetes as a Service for you to use
- Installation and Management of the Kubernetes Master components (aka. Control plane) is handled by Amazon
  - Hosts the control plane in multiple availability zones
  - Fully compatible with standard Kubernetes
  - Upgrades and security patches handled by Amazon
- Cost is $0.20 per hour for control plane
  - Plus standard costs for worker nodes needed to run pods



Reference Architecture

Main Route Table

Destination	Target
10.0.0.0/16	local
pl-68a54001 (com.amazonaws.us-west-2.s3)	vpc-id
0.0.0.0/0	nat-gateway-id

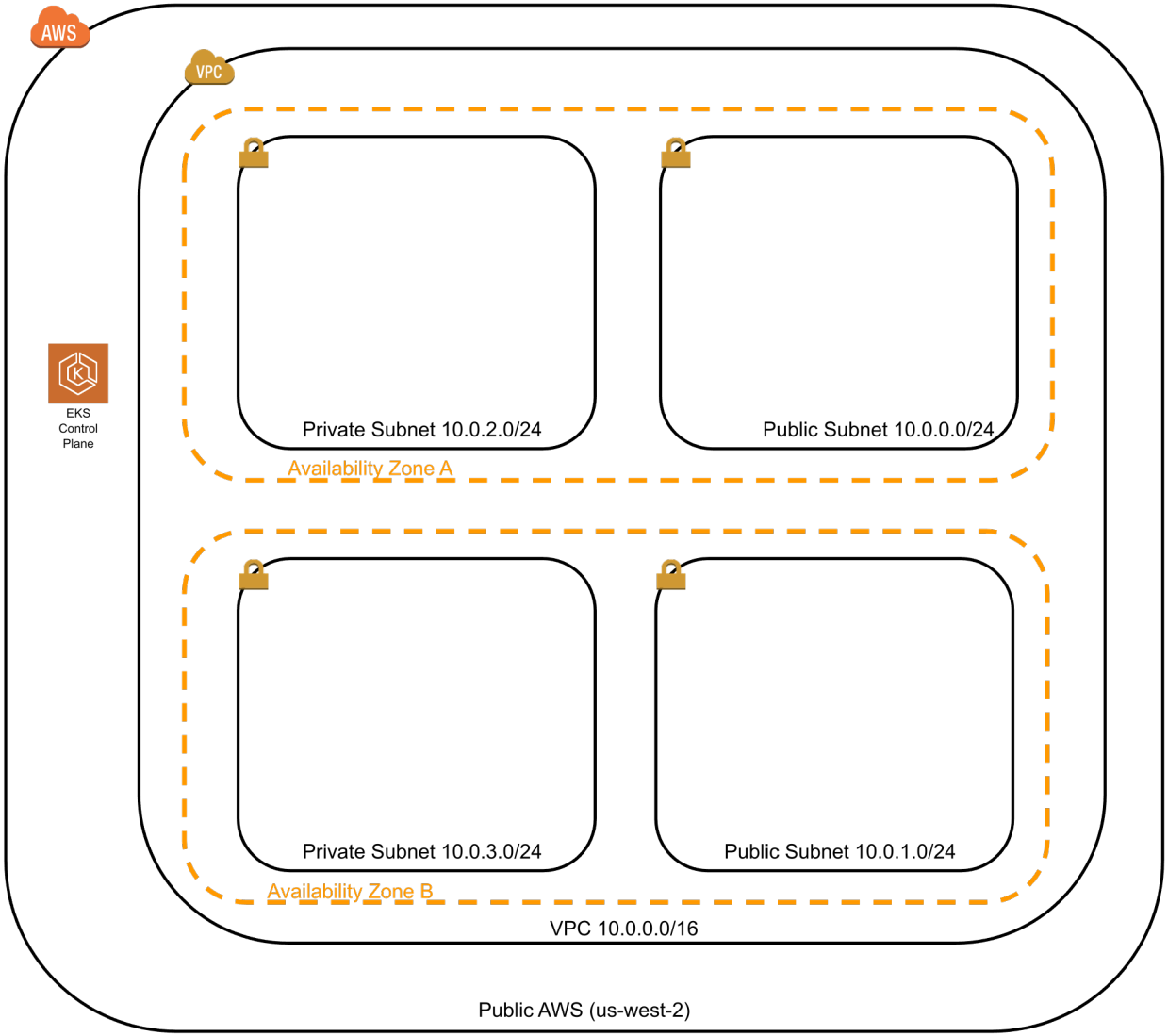


Custom Route Table

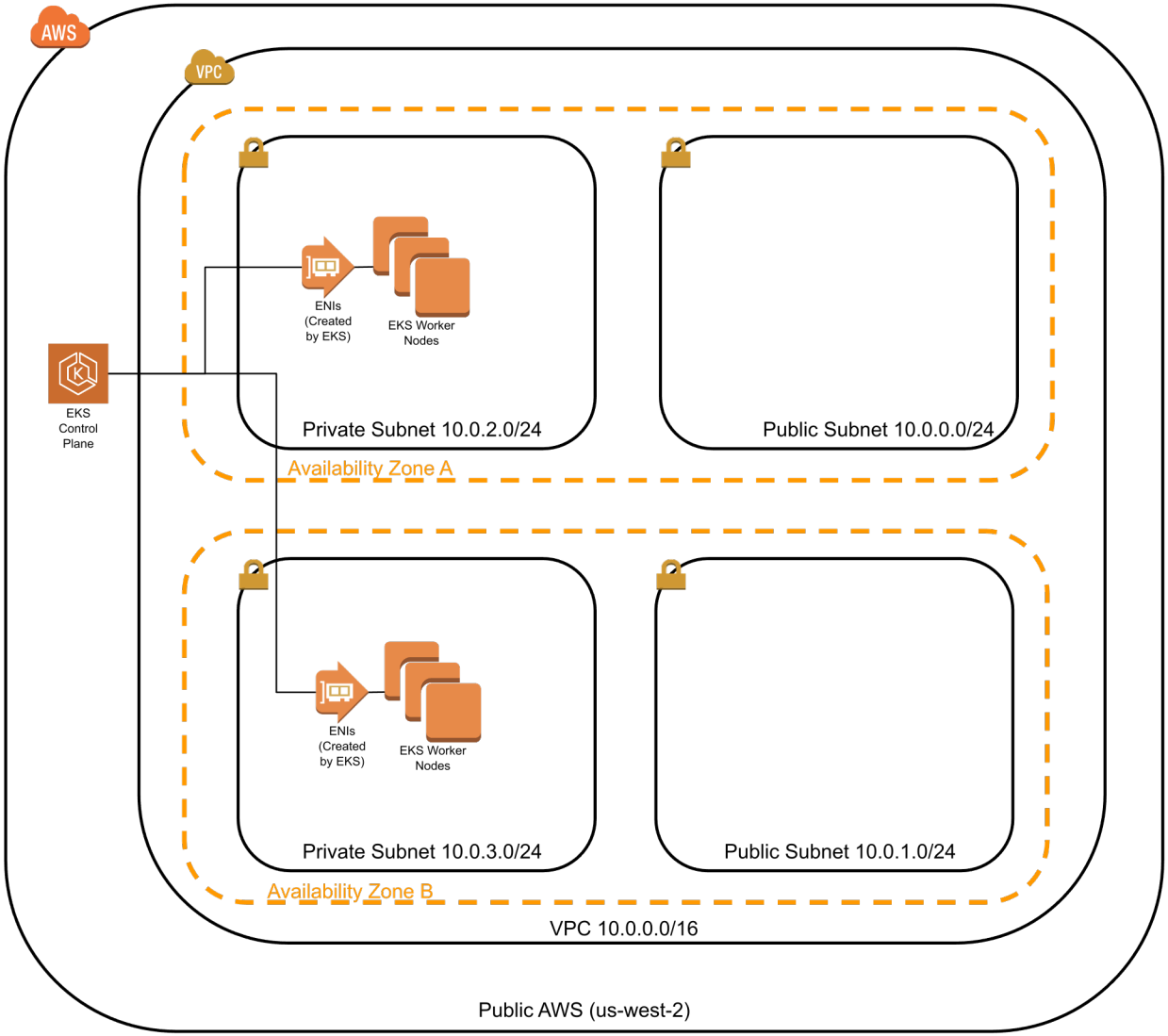
Destination	Target
10.0.0.0/16	local
0.0.0.0/0	igw-id

Reference Architecture (cont.)

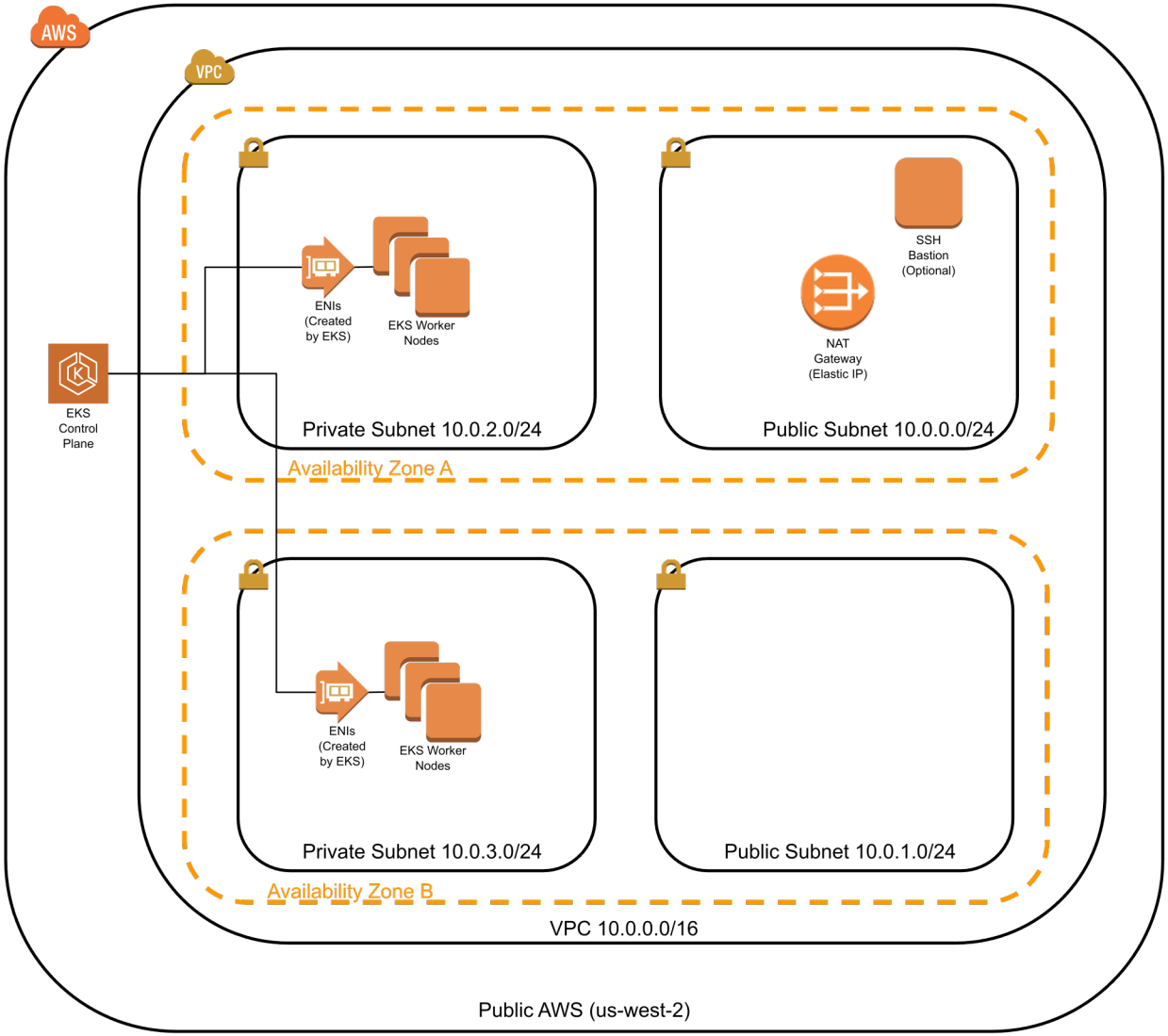
Reference Architecture (cont.)



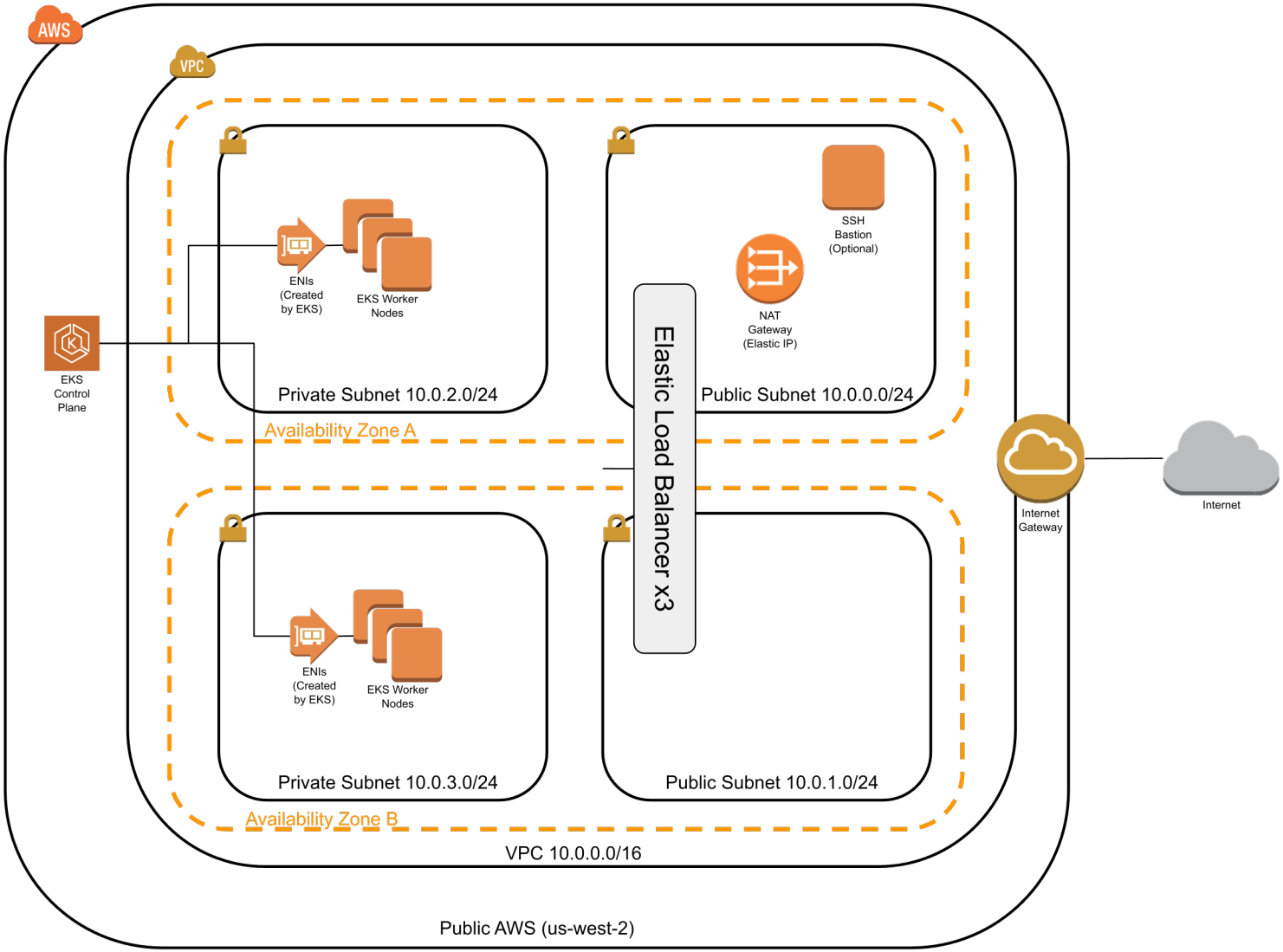
Reference Architecture (cont.)



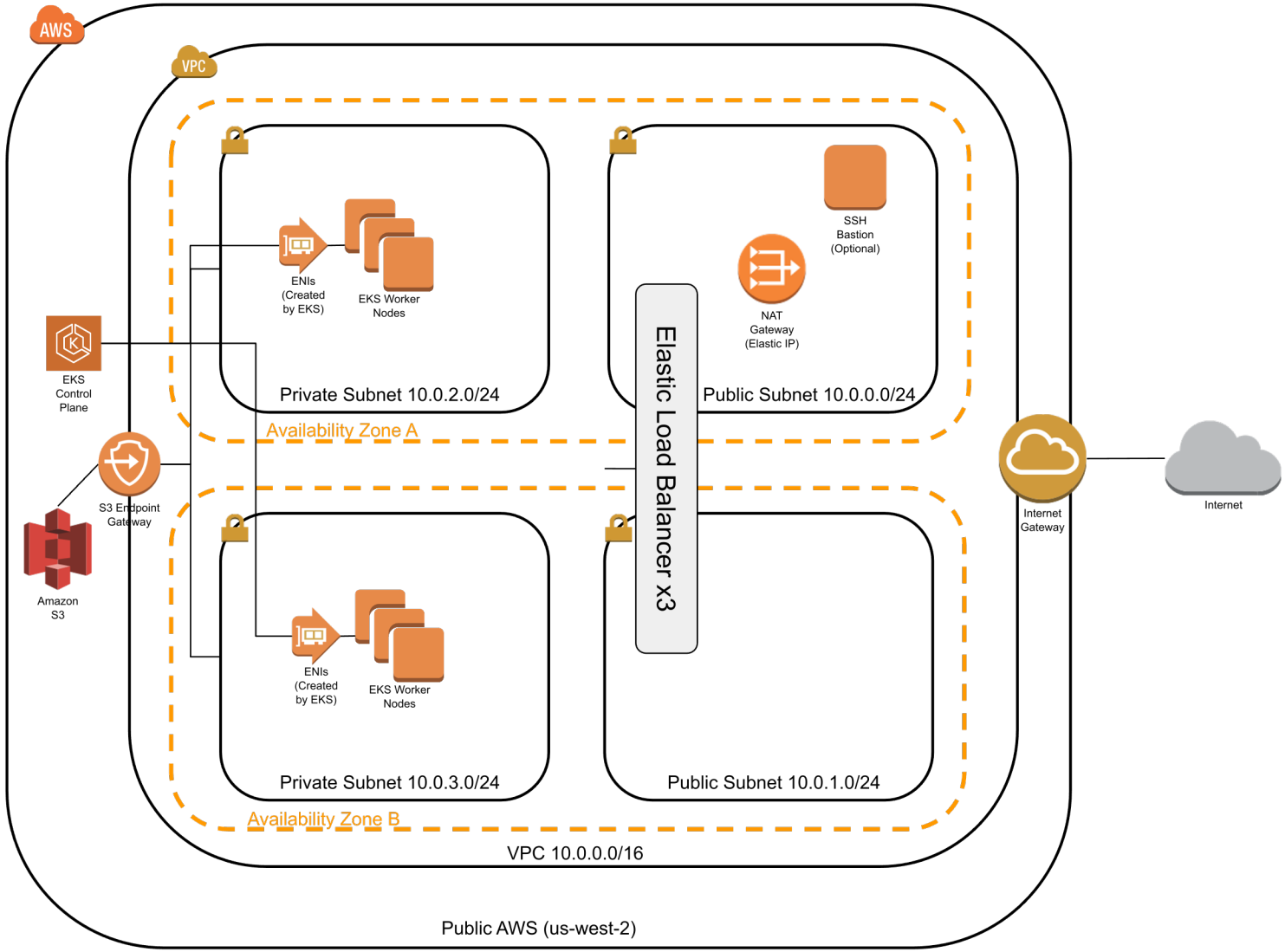
Reference Architecture (cont.)



Reference Architecture (cont.)



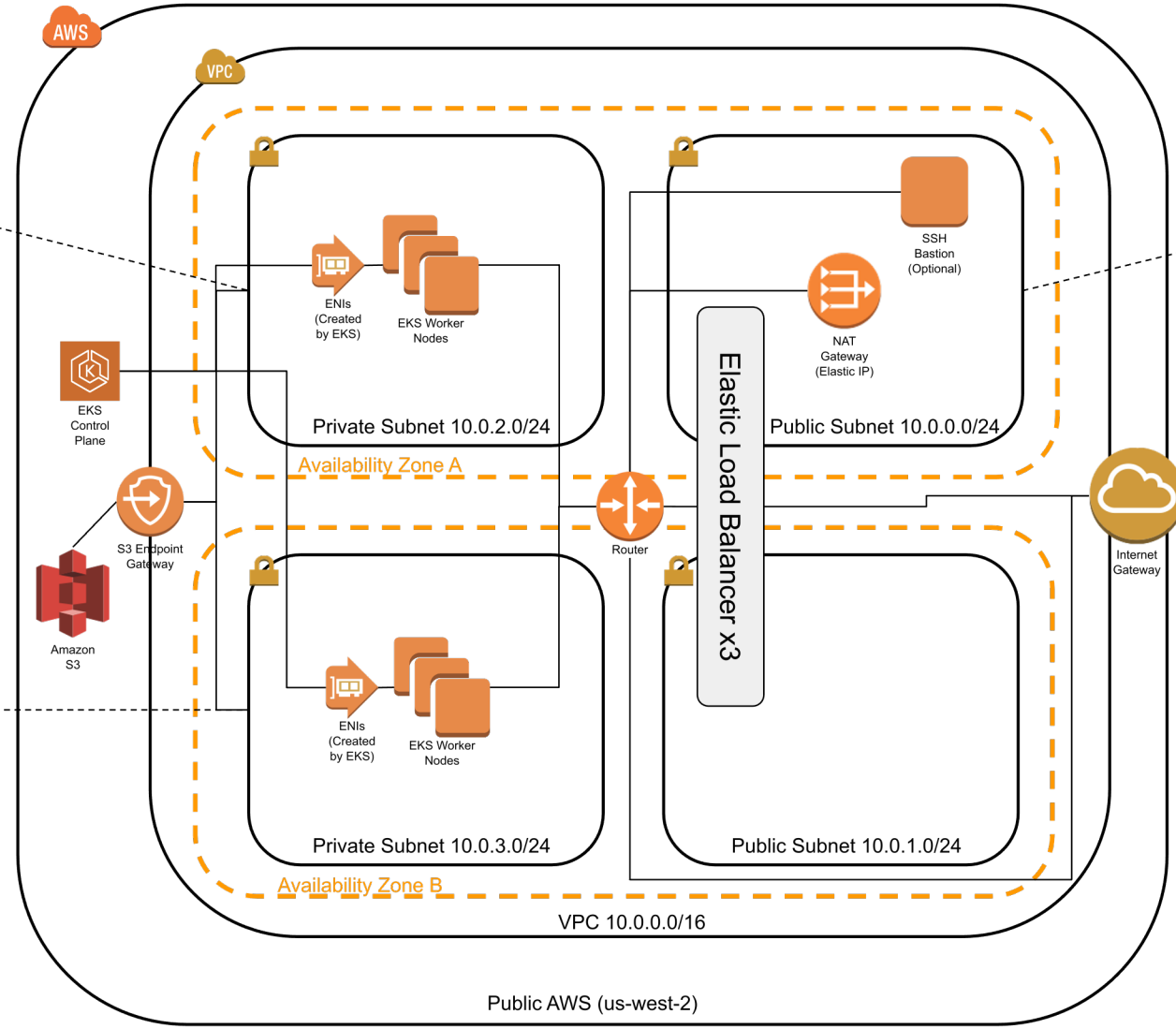
Reference Architecture (cont.)



Reference Architecture (cont.)

Main Route Table

Destination	Target
10.0.0.0/16	local
pl-68a54001 (com.amazonaws.us-west-2.s3)	vpc-id
0.0.0.0/0	nat-gateway-id

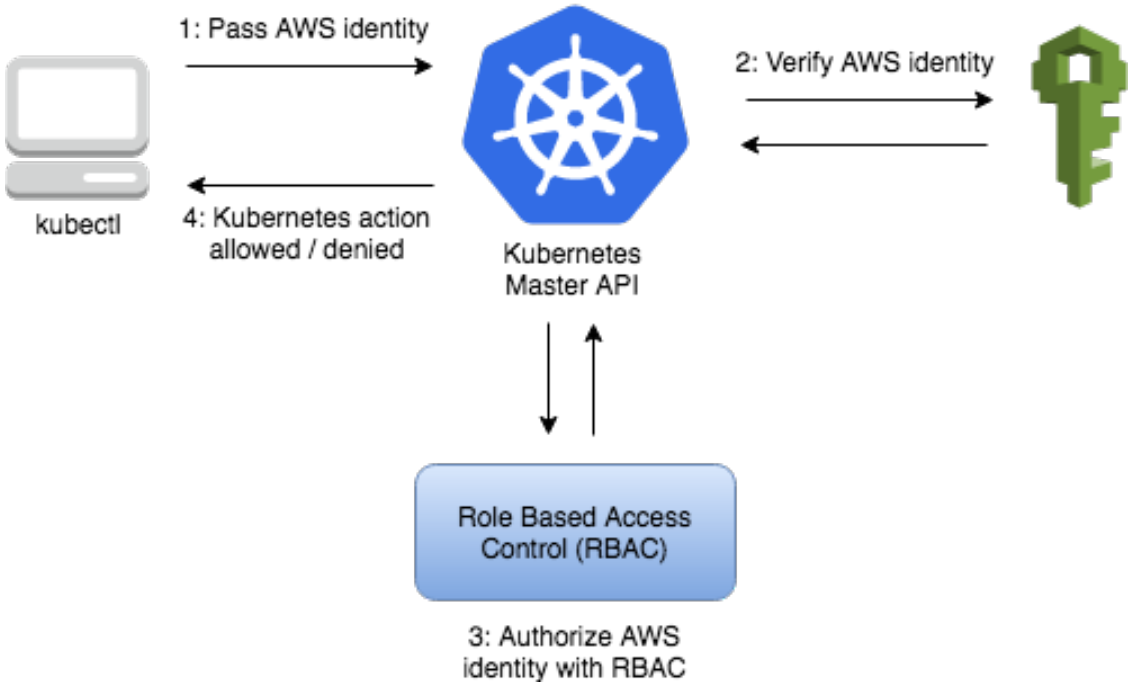


Custom Route Table

Destination	Target
10.0.0.0/16	local
0.0.0.0/0	igw-id

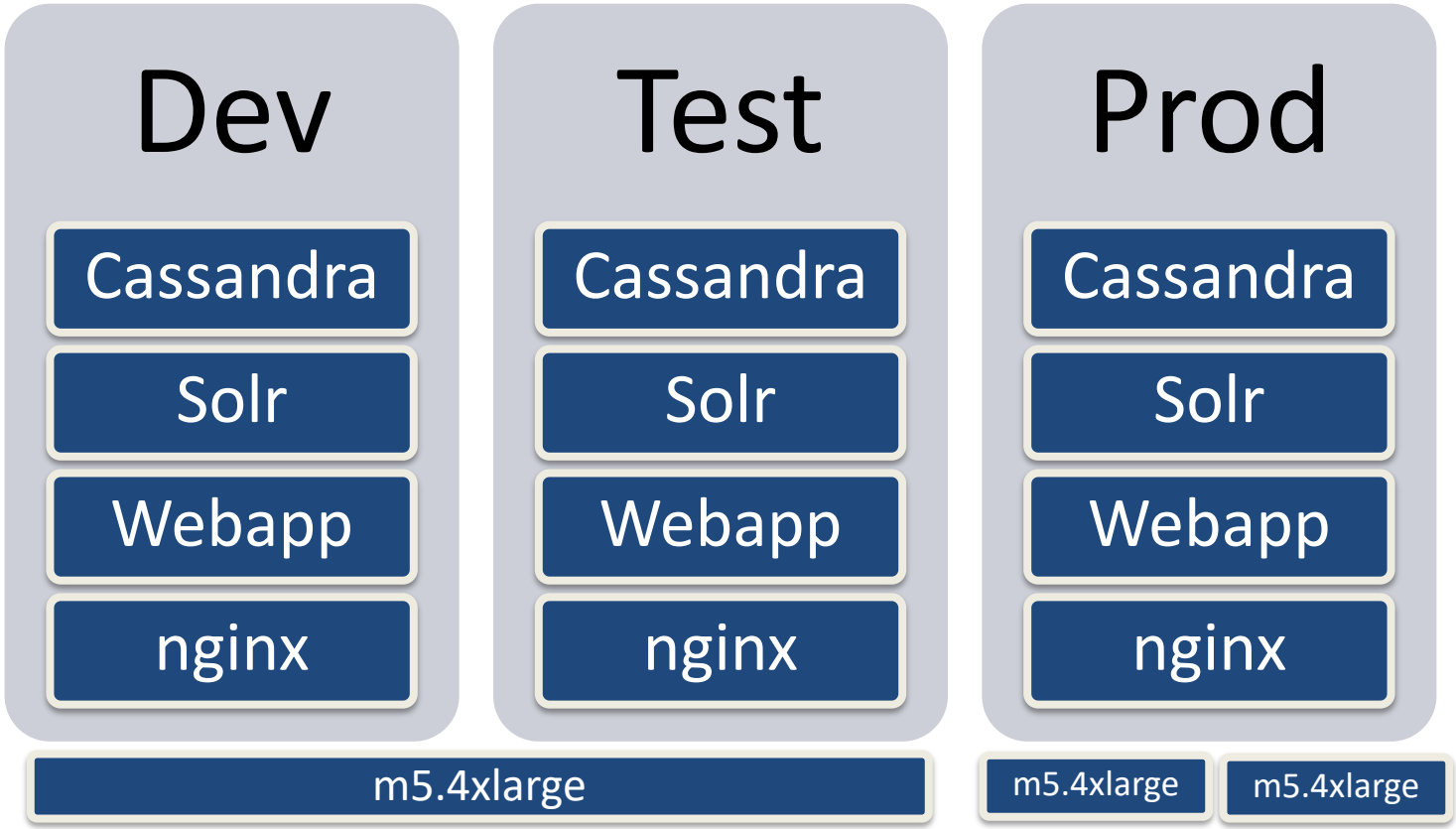
Authentication and Authorization

- Authentication in EKS is handled by IAM
  - Tool called AWS IAM Authenticator for Kubernetes
  - Authenticator can map IAM Users to Kubernetes RBAC Groups
- Authorization is handled by Kubernetes Role Based Access Control (RBAC)
- Groups control who has what access to the resources in the cluster



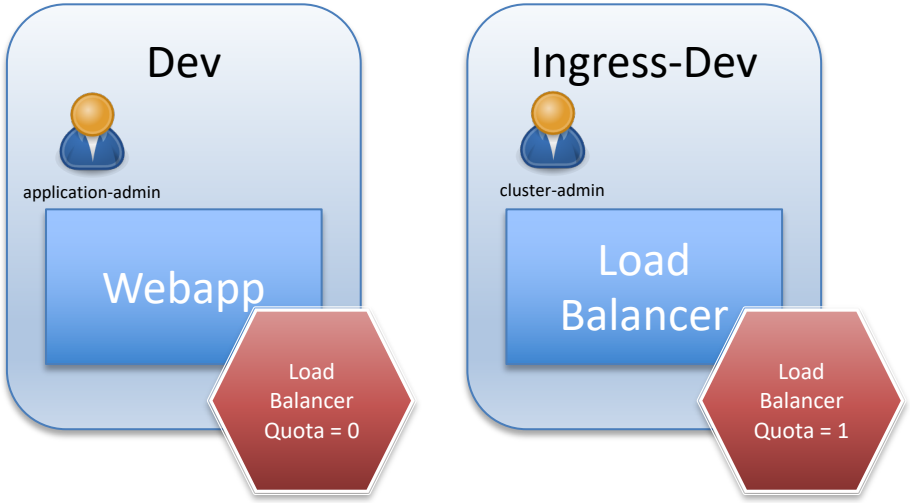
Separation of Environments

- Namespaces provide a mechanism for demarcating environments
- Resources in one namespace cannot interact with resources of another
- Quotas and Security can be managed by namespaces
- Default Taints and Tolerations can “reserve” nodes for use by a particular namespace

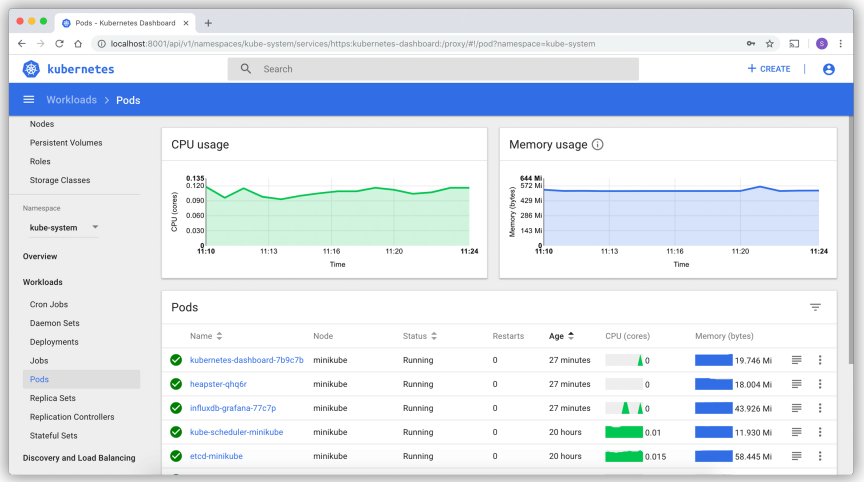


Resource Quotas

- Quotas provide a mechanism to control amount of resources consumed
- Quotas can be defined
  - Cluster level
  - Namespace level
- Quota Types
  - Compute
  - Storage
  - Object count
- Very useful when combined with RBAC
  - For example, limiting who can deploy public-facing ELB

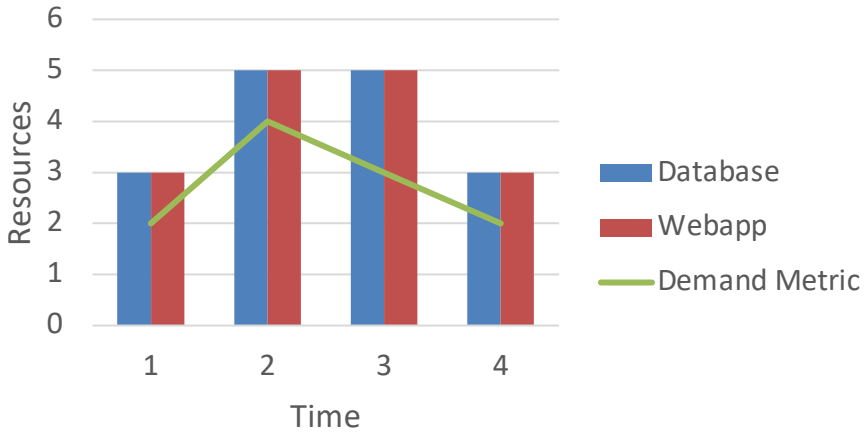


Scaling and Operations



- Cluster Autoscaling
  - Adding new nodes as demand increases
  - Supported by major cloud providers
- Pod Autoscaling
  - HorizontalPodAutoscaler can be used to increase replicas based on resource consumption
  - Based on per-pod resource metric (like CPU utilization)
  - Can be extended with custom metrics

- Web Dashboard gives quick insight into your cluster
- Quickly see resource usage and application failures



Resources

- https://sealevel.nasa.gov/
- https://sdap.apache.org/
- https://incubator-sdap-nexus.readthedocs.io/en/latest/
- https://github.com/apache/incubator-sdap-nexus
- https://towardsdatascience.com/key-kubernetes-concepts-62939f4bc08e
- https://landscape.cncf.io/

Questions?